

Bulletin of Mathematics

Vol. 03, No. 01 (2011), pp. 25–38.

MEMBANDING METODE MULTIPLICATIVE SCATTER CORRECTION (MSC) DAN STANDARD NORMAL VARIATE (SNV) PADA MODEL KALIBRASI PEUBAH GANDA

ARNITA DAN SUTARMAN

Abstract. *Pre processing FTIR spectra prior to the Principle Component Regression (PCR) Analysis is commonly used in multivariate calibration. Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) is one of scatter correction methods the of pre processing. This spectra is corrected by using slope and intercept of regression equation between each spectra of the average of any samples. Calibration PCR model with corrected spectra with MSC showed significant improvement in producing model estimation (RMSEP scatter corrected = 0,1096 versus RMSEP with scatter corrected SNV = 0,1427). Corrected spectra with MSC removed almost 24% of the variance and substantially decreased standard error of prediction of the spectra*

1. PENDAHULUAN

Dewasa ini banyak dikembangkan model kalibrasi peubah ganda untuk mengatasi permasalahan waktu dan biaya [13] Pada percobaan-percobaan kimia, penelitian kandungan (konsentrasi) senyawa aktif suatu tanaman memerlukan tahapan yang panjang dan rumit. Selain itu tidak sedikit biaya yang dibutuhkan untuk pengadaan bahan selama persiapan sampel

Received 07-03-2010, Accepted 15-09-2010.

2000 Mathematics Subject Classification: 62J05, 60B15

Key words and Phrases: Fourier Transform Infra Red, multiplicative signal correction, standard normal variate, principle component regression.

sampai pada pengukuran menggunakan HPLC (High Performance Liquid Chromatography), sehingga diperoleh konsentrasi senyawa aktif yang diinginkan. Pengukuran lain yang berbentuk kualitatif dan relatif lebih sederhana karena prosesnya yang cepat dan membutuhkan biaya yang murah yaitu FTIR (Fourier Transform Infrared). Namun keluaran yang dihasilkan FTIR ini hanya berupa spektrum besarnya nilai serapan saat sampel disinari infra merah. Oleh karena itu dikembangkanlah model kalibrasi peubah ganda yang menggambarkan bentuk hubungan antara satuan pengukuran yang dilakukan dengan cara yang rumit dan relative mahal yaitu HPLC. Data keluaran FTIR adalah data berbentuk spektrum yang menunjukkan nilai serapan saat contoh disinari infra merah. Data ini merupakan data kontinu terhadap bilangan gelombang. Nilai pengukuran pada suatu bilangan gelombang dipengaruhi oleh nilai-nilai pada bilangan gelombang sebelumnya, sehingga permasalahan korelasi menjadi salah satu kendala dalam pemodelan kalibrasi. Selain itu, permasalahan dimana jumlah amatan yang lebih kecil dari jumlah variabel ($n < p$) juga merupakan kendala lainnya yang harus diatasi, karena hal ini dapat menghasilkan penyelesaian yang tidak stabil jika menggunakan regresi peubah ganda [4, 14].

Terdapat beberapa metode yang sering digunakan untuk mengatasi permasalahan-permasalahan di atas, diantaranya adalah Regresi Komponen Utama (RKU). Cara kerja metode RKU adalah matriks data dekomposisi sampai diperoleh r komponen-komponen utama yang bertindak sebagai peubah-peubah baru yang saling orthogonal dengan jumlah komponen yang lebih kecil dan jumlah peubah asal ($r \ll p$) dan jumlah amatan ($n > r$).

Tujuan dari penelitian ini adalah mengevaluasi model kalibrasi yang diperoleh melalui koreksi pencaran. Karena itu dalam penelitian ini akan dilakukan koreksi pencaran terhadap nilai kemiringan dan intersep dari model regresi tiap sampel. Membandingkan dua metode koreksi pencaran yaitu Multiplikatif Scatter Correction (MSC) dan Standard Normal Variance (SNV). Dilanjutkan dengan membangun model kalibrasi peubah ganda terhadap data yang sudah dikoreksi tersebut.

2. DATA DAN METODE

Data yang digunakan dalam penelitian ini adalah data simulasi berupa data yang dibangkit melalui bantuan software statistika.

Langkah-langkah analisis data yang akan dilakukan dalam penelitian ini adalah:

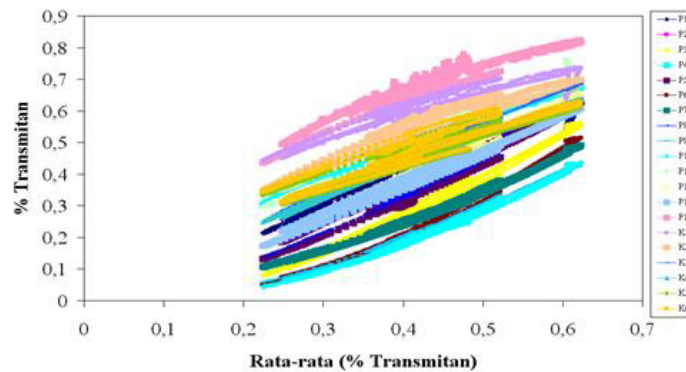
1. Spektrum semua sampel diplot terhadap rata-rata untuk melihat pola garis lurus yang dibentuk oleh spektrum tiap contoh tersebut.

2. Membentuk model regresi linier $x_i = \beta_{0i} + \beta_{1i}\bar{x}_j + e_i$ dan menduga parameter menggunakan metode kuadrat terkecil.
3. Setelah parameter β diperoleh kemudian digunakan untuk mentransformasi data asli dengan mengurangi nilai amatan x_{ij} terhadap konstanta $\hat{\beta}_{0i}$ dan membagi hasil pengurangan terhadap $\hat{\beta}_{1i}$, sehingga diperoleh bentuk transformasinya $x_{ij}^* = (x_{ij} - \hat{\beta}_{0i})/\hat{\beta}_{1i}\bar{x}_j$.
4. Spektrum semua sampel ditransformasi dengan cara mengurangi tiap nilai amatan pada tiap spektrum terhadap rata-rata tiap sampel spektrum tersebut.
5. Membagi data yang sudah ditransformasi menjadi dua bagian untuk tujuan pembentukan model kalibrasi peubah ganda dan validasi model. Pada penelitian ini data yang digunakan terdiri dari dua puluh contoh. Lima belas data digunakan untuk pengembangan model kalibrasi dan lima data selebihnya digunakan untuk validasi model. Pemilihan contoh yang digunakan untuk pembentukan model dan validasi model dilakukan dengan pengacakan. Data bagian pertama digunakan untuk membentuk model kalibrasi peubah ganda.
6. menggunakan Analisis Komponen Utama (AKU) pada masing-masing hasil transformasi baik MSC maupun NSV
7. Komponen-komponen utama yang diperoleh ditambah peubah boneka selanjutnya diregresikan terhadap nilai pengukuran HPLC
8. Data bagian kedua digunakan untuk validasi model kalibrasi peubah ganda sehingga diperoleh nilai konsentrasi dugaan.
9. Evaluasi kebaikan model dengan melihat beberapa kriteria yaitu RMSEP (it Root Mean Square Error Prediction), MSEP (Mean Square Error Prediction) dan SEP (*Standard Error Prediction*).
10. Memodelkan data yang tidak dikoreksi dan membandingkannya terhadap model pada data yang dikoreksi baik yang menggunakan MSC maupun NSV

Koreksi Pencaran

Sebelum membentuk model kalibrasi, data yang sudah diacak kemudian dikoreksi terhadap nilai kemiringan dan intersep dari model regresi masing-masing contoh. Pengaruh koreksi pencaran yang telah dilakukan dapat dilihat dengan membandingkan plot regresi dan spektrum keseluruhan

contoh pada data yang sudah dikoreksi menggunakan metode MSC dan data yang dikoreksi menggunakan SNV. Pada kasus kalibrasi peubah ganda plot penumpangtindihan semua spektra mampu memberikan informasi tentang gejala keanehan data seperti penciran dan identifikasi daerah noise.



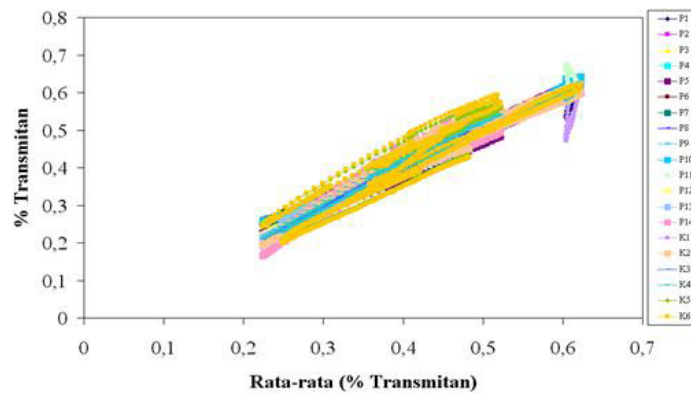
Gambar 1: Plot persen transmittan seluruh contoh yang dikoreksi terhadap terhadap rata-rata pada data yang dikoreksi menggunakan metode SNV.

Pada Gambar 1 terlihat bahwa nilai spektra contoh-contoh tersebut mendekati garis lurus, namun garis regresi antar contoh berbeda pada nilai kemiringan dan intersepnya. Perbedaan tersebutlah yang diidentifikasi sebagai perbedaan yang diakibatkan adanya pengaruh pencaran yang menyebabkan informasi yang ada pada tiap contoh juga berbeda. Sedangkan penyimpangan contoh-contoh dari garis regresinya diinterpretasikan sebagai informasi kimia pada spektra.

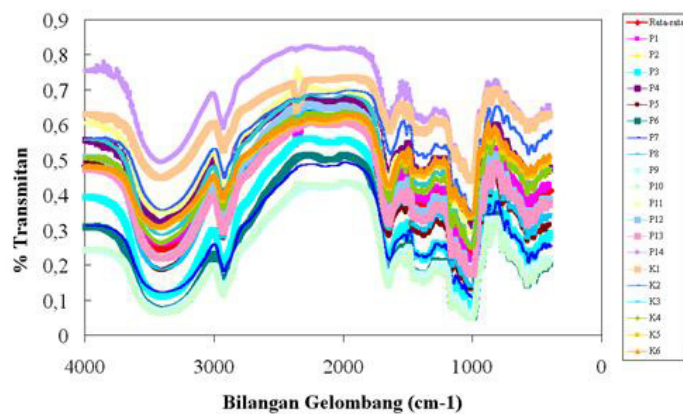
Setelah pencaran dikoreksi menggunakan metode MSC ternyata garis regresi tiap contoh berada pada kemiringan dan intersep yang hampir sama. Hal ini dapat diartikan bahwa informasi yang diberikan pada masing-masing spektrum sudah relatif sama.

Pengaruh transformasi pada data dapat dilihat pada Gambar 3 dan 4, terlihat bahwa spektrum pada data yang dikoreksi menggunakan MSC, pencarannya mampu menjadikan pola dan posisi pencaran lebih mendekati pola dan posisi pencaran rujukannya, dalam hal ini adalah pencaran rata-rata seluruh contoh yang terletak pada kisaran nilai persen transmittan 0,4 sampai 0,5.

Pada Gambar 3 dan 4, dapat dikatakan bahwa data yang dikoreksi menggunakan MSC mampu mengeliminasi keragaman antar spektra yang didominasi oleh pengaruh aditif dan multiplikatif. Plot antar contoh terlihat



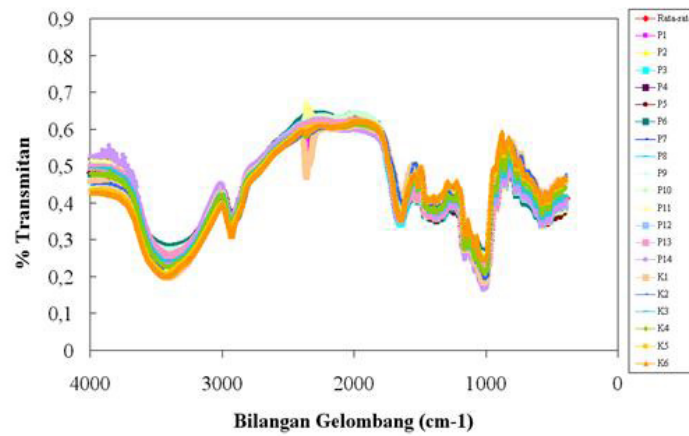
Gambar 2: Plot persen transmittansi seluruh contoh terhadap rata-ratanya pada data yang dikoreksi menggunakan metode MSC.



Gambar 3: Plot spektrum seluruh contoh pada data yang dikoreksi menggunakan SNV.

lebih rapat, kemiringan tiap spektrum lebih kecil dan bergeser mendekati rata-ratanya, sehingga informasi yang diberikan spektrum antar contoh sudah relatif sama.

Data keluaran spektrum merupakan bilangan gelombang yang bersifat kontinu, oleh karenanya data tersebut berautokorelasi antara satu pengamatan terhadap pengamatan sebelumnya. Namun dalam koreksi pencaran, memodelkan spektrum pada satu atau beberapa contoh terhadap rata-rata-



Gambar 4: Plot spektrum seluruh contoh pada data yang dikoreksi menggunakan MSC.

nya tidak memerlukan asumsi seperti dalam pembentukan model regresi pada umumnya, karena dalam kasus koreksi pencaran yang diperlukan hanya parameter sebagai faktor koreksi. Pemodelan ini hanya bertujuan untuk melihat besar pergeseran spektra baik aditif atau multiplikatif terhadap rata-ratanya. Sedangkan galat pada model disesuaikan pada semua pengaruh spektrum yang tidak dimodelkan oleh konstanta aditif dan multiplikatif. Menurut Pyndick dan Rubinfeld [15], autokorelasi tidak mempengaruhi ketidakbiasan pendugaan parameter tetapi akan menghasilkan ragam yang lebih kecil dari ragam yang sebenarnya.

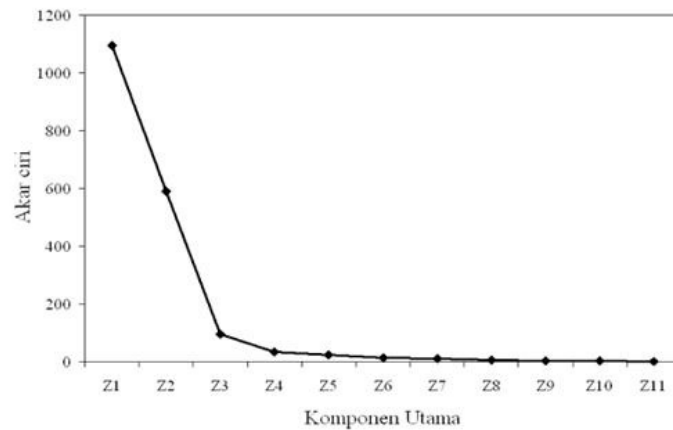
Pembentukan skor komponen utama Data yang sudah dikoreksi terhadap koefisien regresi dibagi menjadi dua bagian dengan porsi yang berbeda untuk tujuan pembentukan model kalibrasi dan validasi. Hasil dari analisis komponen utama pada lima belas sampel diperoleh sebelas skor komponen pertama yang saling bebas, artinya korelasi antar komponen utama telah teratasi. Pemilihan sebelas komponen utama pertama ini didasari pada nilai kumulatif keragaman yang diberikan oleh komponen-komponen utama tersebut mencapai seratus persen. Nilai akar ciri, proporsi keragaman dan kumulatif keragaman pada sebelas komponen utama pertama yang telah diurutkan berdasarkan keragaman terbesar hingga terkecil disajikan pada Tabel 1.

Tabel 1 memperlihatkan bahwa tiga komponen utama pertama merupakan komponen utama yang menjelaskan keragaman terbesar dengan pro-

Tabel 1: Nilai akar ciri (λ_i) pada sebelas komponen utama pertama

Komponen Utama	Akar ciri	Proporsi Keragaman (%)	Kumulatif Keragaman (%)
Z1	1094,5	58,7	58,7
Z2	589,1	31,6	90,2
Z3	95,9	5,1	95,4
Z4	34,3	1,8	97,2
Z5	22,1	1,2	98,4
Z6	11,6	0,6	99,0
Z7	9	0,5	99,5
Z8	4,2	0,2	99,7
Z9	2,2	0,1	99,8
Z10	1,5	0,1	99,9
Z11	0,9	0,1	100,0

porsi keragaman masing-masing komponen utama adalah 58,7%, 31,6% dan 5,1%. Keragaman mencapai seratus persen pada komponen utama ke sebelas dengan proporsi keragaman 1%. Penurunan proporsi keragaman dari sebelas komponen utama tersebut dapat digambarkan pada plot penurunan akar ciri pada Gambar 5.



Gambar 5: Plot penurunan nilai akar ciri pada sebelas komponen utama pertama.

Regresi Komponen Utama

Sebelas skor komponen utama yang diperoleh melalui analisis komponen utama dijadikan peubah baru yang akan digunakan membentuk model kalibrasi. Umumnya komponen utama diurutkan berdasarkan nilai keragaman terbesar hingga terkecil. Beberapa komponen utama terakhir sering dieliminasi tanpa merasa kehilangan suatu informasi yang penting karena dianggap memberikan sedikit sekali keragaman. Kesebelas komponen utama tersebut ditambah satu peubah boneka yang menjelaskan masa simpan serbuk tanaman dimodelkan dengan membentuk persamaan regresi berganda terhadap nilai Y . Prosedur pemilihan komponen utama yang dilakukan adalah dengan memasukkan setiap komponen utama berdasarkan akar cirinya ke dalam model. Evaluasi model terbaik pada himpunan komponen utama yang terbentuk adalah dengan melihat nilai RMSEP minimum. Pada Lampiran 1 disajikan sebelas model yang disusun dengan mencobakan setiap komponen utama masuk ke dalam model.

Dari sepuluh model yang diperoleh dari evaluasi model akar ciri memperlihatkan bahwa model keempat adalah model terbaik karena memiliki nilai RMSEP paling minimum yaitu 0,10961. Model tersebut disusun oleh empat komponen utama pertama ditambah satu peubah boneka yang memberi sumbangan keragaman terbesar dengan kumulatif keragaman sebesar 97,2

Tabel 2 menunjukkan bahwa empat komponen utama pertama berpengaruh nyata pada taraf $\alpha = 5\%$ terhadap nilai Y . Begitu juga dengan peubah boneka D (lama penyimpanan). Selisih rata-rata antara nilai Y dengan masa simpan kurang dari 3 bulan dibandingkan dengan masa simpan lebih dari 3 bulan sebesar 0,2811. Jika dilihat dari nilai-nilai Variance Inflation Factor (VIF), komponen-komponen utama tersebut mempunyai nilai VIF kurang dari 10, hal ini mengindikasikan bahwa korelasi antar komponen utama sudah teratasi. Jika ditulis dalam bentuk persamaan regresi peubah ganda maka model tersebut adalah :

$$\hat{y} = 0,781 - 0,0045Z1 - 0,0021Z2 - 0,0094Z3 - 0,0113Z4 + 0,2811D$$

Uji kesesuaian model ditunjukkan pada tabel analisis ragam pada Tabel 3. Berdasarkan nilai- p pada Lack of Fit sebesar 0,944 ($\alpha = 5\%$) maka model telah sesuai

Validasi model

Setelah model kalibrasi dibentuk menggunakan data bagian pertama, kemudian model divalidasi menggunakan data bagian kedua, dalam kasus ini data yang digunakan untuk validasi adalah data lima contoh serbuk yang diperoleh dengan pengacakan. Model kalibrasi yang dibentuk dalam kom-

Tabel 2: Model kalibrasi peubah ganda dengan prosedur seleksi evaluasi akar ciri

Peubah penjelas	Koef	SE Koef	T-Hit	Nilai-p	VIF
Konstanta	0,7809	0,0444	17,61	0	
Z1	0,0045	0,0015	-3,04	0,014	3,8
Z2	0,0021	0,0011	-1,87	0,094	1,1
Z3	0,0094	0,0027	-3,48	0,007	1,1
Z4	0,0113	0,0052	-2,16	0,059	1,5
D	0,2811	0,1108	2,54	0,032	4,5

Tabel 3: Uji kesesuaian model kalibrasi peubah ganda

Sumber Keragaman	DB	Jumlah Kuadrat	Kuadrat Tengah	<i>F</i> -Hit	Nilai- <i>p</i>
Regresi	5	1,42467	0,28493	31,54	0
Galat	9	0,08131	0,00903		
Lack of Fit	8	0,05006	0,00626	0,2	0,944
Pure Error	1	0,03125	0,03125		
Total	14	1,50597			

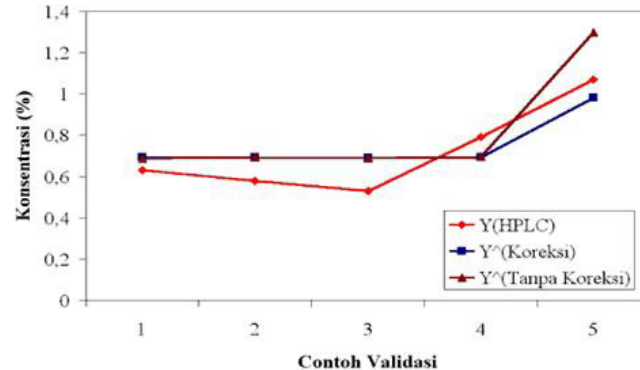
Tabel 4: Nilai konsentrasi dugaan (\hat{Y}) pada contoh validasi

Y	(Koreksi MSC)	(Koreksi SNV)
0,63	0,695	0,688
0,58	0,693	0,695
0,53	0,691	0,689
0,79	0,695	0,696
1,07	0,980	1,297

ponen utama kemudian dikembalikan menjadi bentuk perkalian koefisien komponen utama dan peubah asalnya. Dari proses validasi, diperoleh nilai dugaan dari model yang mencerminkan bentuk peubah asalnya.

Nilai dugaan yang diperoleh dari model kalibrasi peubah ganda dengan menggunakan contoh validasi dapat dilihat pada Tabel 4.

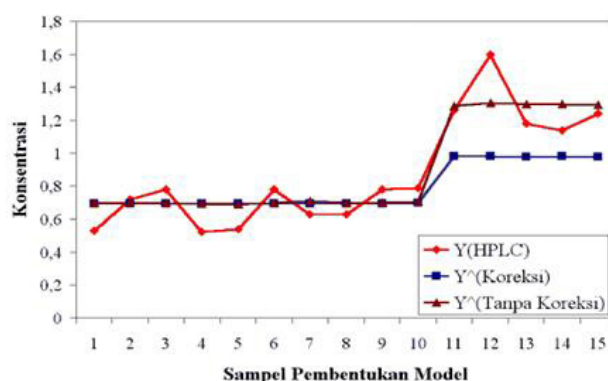
Jika dibandingkan nilai Y sebenarnya terhadap nilai Y dugaan pada data dikoreksi MSC dan dikoreksi SNV, diperoleh selisih antara kedua nilai dugaan tidak begitu besar kecuali pada contoh terakhir. Perbandingan ketiga nilai Y (konsentrasi) ini dapat dilihat pada Gambar 6.


Gambar 6: Plot nilai Y dan nilai Y dugaannya pada contoh validasi.

Pada Gambar 6 terlihat bahwa nilai Y dugaan pada data yang dikoreksi MSC mendekati nilai Y duga pada data yang dikoreksi SNV kecuali pada contoh terakhir. Perbandingan kedua nilai Y dugaan terhadap nilai Y hasil pengukuran juga cukup dekat kecuali pada data terakhir, data yang

dikoreksi MSC lebih dekat dibanding data yang dikoreksi SNV.

Perbandingan nilai Y hasil pengukuran dengan nilai dugaan data yang dikoreksi MSC dan dikoreksi SNV pada contoh yang digunakan untuk membentuk model kalibrasi dapat dilihat pada Gambar 7. Dapat dijelaskan bahwa secara keseluruhan nilai dugaan yang diperoleh dengan mengoreksi dan tanpa mengoreksi pencaran pada contoh pembentukan model cukup dekat kecuali pada lima contoh terakhir yang merupakan contoh dengan masa penyimpanan sebentar. Namun nilai dugaan yang diperoleh dengan mengoreksi pencaran mampu menurunkan galat jika dievaluasi dari nilai MSE, RMSEP dan SEP kedua model tersebut. Dapat dikatakan bahwa proses pembentukan model kalibrasi peubah ganda jika datanya dikoreksi mampu memberikan model yang lebih baik dibanding model kalibrasi jika datanya tidak dikoreksi. Perbandingan tersebut dapat dilihat pada beberapa kriteria kebaikan model pada Tabel 5.



Gambar 7: nilai Y dan nilai Y dugaannya pada contoh pembentukan model.

Model kalibrasi yang dilakukan koreksi pencaran pada contoh pengujian mampu menurunkan nilai MSE, RMSEP, dan SEP masing-masing sebesar 41,24%, 23,35% dan 2,82%.

4. KESIMPULAN

Koreksi pencaran pada data senyawa aktif gingerol serbuk rimpang jahe penting dilakukan untuk memperoleh model yang lebih akurat. Koreksi pencaran juga dapat menjadikan pola dan posisi spektrum tiap contoh lebih mendekati rujukannya dalam hal ini spektrum rata-rata keseluruhan contoh.

Tabel 5: Perbandingan model kalibrasi pada pencaran yang dikoreksi MSC dan dikoreksi SNV

Satuan Pengukuran	\hat{Y} (Koreksi MSC)	\hat{Y} (Koreksi SNV)	% Penurunan Galat
MSEP	0,0601	0,1022	41,24
RMSEP	0,1096	0,1430	23,35
Deviasi	0,0306	0,0930	67,08
SEP	0,0170	0,0175	2,82

Hal ini mengakibatkan informasi yang diberikan spektrum tiap contoh relatif sama.

Pemodelan kalibrasi dilakukan dengan meregresikan komponen-komponen utama ditambah peubah boneka yang menjelaskan masa simpan serbuk rimpang jahe terhadap nilai pengukuran HPLC. Komponen-komponen utama yang terbentuk diperoleh dengan metode komponen utama, dimana metode tersebut mampu mengatasi permasalahan data dimensi besar dan korelasi yang tinggi antar peubah bebasnya. Model kalibrasi yang dibentuk dari data yang dikoreksi pencarannya mampu memberikan nilai RMSEP, MSEP dan SEP yang lebih kecil dibanding model kalibrasi yang dibentuk dari data yang tidak dikoreksi pencarannya.

Daftar Pustaka

- [1] Atok RM, Notodiputro KA. 2004. Metode NN (Neural Network) dengan Principle Component sebagai Pre processing pada Data. *Proceeding Seminar Nasional Statistika*, IPB, Bogor.
- [2] De Maesschalck R et al. 1999. The Development of Calibration Models for Spectroscopic Data Using PCR. *Internet Journal of Chemistry* 2.
- [3] Draper N, Smith H. 1981. *Analisis Regresi Terapan. Ed ke-2. Bambang Sumantri, penerjemah*. Jakarta: Gramedia Pustaka Utama. Terjemahan dari: Applied Regression Analysis.
- [4] Erfiani, Notodiputro KA. 2004. Penggunaan Winbugs untuk Pendugaan Model Kalibrasi. *Proceeding Seminar Nasional Statistika*, IPB, Bogor.

- [5] Frank IE, Friedman JH. 1993. A statistical View of Some Chemometrics Regression Tools. *Technometrics* 35(2):109-135.
- [6] Johnson RA, Wicher DW. 2002. *Applied Multivariate Statistical Analysis*. Ed ke-5. NJ, USA: Prentice-Hall, Inc. Upper Saddle River.
- [7] Jolliffe IT. 1986. *Principal Component Analysis*. New York: Springer-Verlag..
- [8] Jorgensen A. 2000. *Clustering Excipient Near Infrared Spectra Using Different Chemometric Methods*. <http://www.google.com>. [13 September 2004]
- [9] Krzanowski WJ. 1990. *Principle of Multivariate Analysis*. New York: Oxford University Press.
- [10] Meyers RA, editor. 2000. Chemometric Methods in Process Analysis. *Encyclopedia of Analytical Chemistry*. Chichester: Jhon Wiley & Sons Ltd. hlm 8145-8167.
- [11] Meyers RH. 1986. *Classical and Modern Regression with Applications*. United States of America: PWS Publishers.
- [12] Naes T, Issakson T, Fearn T, Davies T. 2002. *Multivariate Calibration and Classification*. United Kingdom: NIR Publications Chichester.
- [13] Samp EJ, Sedin D, Foster A. 2003. Enhanced NIR Calibration for Wort Fermentability Using Orthogonal Signal Correction. *Journal of Institute & Guild of Brewing*. 109(1):16-26.
- [14] Sunaryo S, Notodiputro KA. 2004. Reduksi Dimensi Data Spektra dengan Transformasi Fourier dan Wavelet. *Proceeding Seminar Nasional Statistika*, IPB, Bogor.
- [15] Pindyck RS, Rubinfeld DL. 1998. *Econometric Models and Economic Forecasts*. Ed ke-4. Singapore: McGraw-Hill Companies.
- [16] Weisberg S. 1985. *Applied Linier Regression*. Ed ke-2. New York: John Wiley & Sons.

ARNITA: Jurusan Bahasa Indonesia FBS UNIMED, Jl. Willem Iskandar Medan, Indonesia.

E-mail: arnita22@yahoo.com

SUTARMAN: Departemen Matematika, FMIPA, Universitas Sumatera Utara, Jl.
Bioteknologi No. 1 Kampus USU, Medan 20155, Indonesia.
E-mail: sutarman@usu.ac.id